Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 8556–8561, July 1998
Biochemistry

In search of a mutational hotspot

ZAFER HATAHET*, MEIXIA ZHOU*†, LINDA J. REHA-KRANTZ‡, SCOTT W. MORRICAL*§, AND SUSAN S. WALLACE*¶

*Department of Microbiology and Molecular Genetics and the Markey Center for Molecular Genetics and §Department of Biochemistry, University of Vermont, Burlington, VT 05405; and ‡Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

Edited by Bruce N. Ames, University of California at Berkeley, CA, and approved May 16, 1998 (received for review January 30, 1998)

ABSTRACT ***In vitro*** **selection was used to define sequence contexts that significantly enhanced the mutagenic potential of 7,8-dihydro-8-oxoguanine (8-oxoG). Contexts that simultaneously reduced the efficiency of 8-oxoG cleavage by formamidopyrimidine DNA *N*-glycosylase and increased the efficiency of misincorporating A opposite the lesion by DNA polymerase were isolated from a pool of 4^8 random octanucleotide sequences. Kinetic analysis showed that the combined effects of poor repair and high miscoding resulted in 10^2- to 10^3-fold increase in the mutagenic potential of 8-oxoG. Furthermore, the isolated sequence contexts correlated strongly with G → T transversion hotspots in spontaneous mutational spectra reported for the *Escherichia coli lacI* and human p53 and factor IX genes. We present an example directly linking the interplay between DNA repair and replication to a "high risk sequence" for base substitution.**

The four base moieties in DNA are constant targets of damage by chemical processes such as depurination, deamination, and oxidation. DNA oxidation is caused by ionizing radiation, chemical agents (1), and, perhaps most importantly, normal metabolism (2). 8-Oxoguanine (8-oxoG) is one of the most prominent base oxidation products (3, 4) and has been implicated in mutagenesis, carcinogenesis, and aging (5). It has been shown to cause G → T transversions *in vivo* (6–9) and to be readily bypassed by DNA polymerases *in vitro*, coding for a nonmutagenic C and to a lesser extent a mutagenic A (10).

The biological consequence of a DNA lesion is determined largely by the outcome of its interactions with the proteins responsible for maintaining the genetic integrity of the cell, namely the replicative and repair machineries. There are two prerequisites for a lesion to cause a mutation; repair enzymes must fail to remove it before the next round of replication takes place, and it must have coding properties different from those of its cognate base. To counter the mutagenic effect of 8-oxoG, *Escherichia coli* has evolved an elaborate repair system consisting of three genes, *mutT*, *mutM*, and *mutY* (7, 11). *mutT* codes for a sanitizing enzyme that converts 8-oxoGTP in the nucleotide pool to 8-oxoGMP, preventing the incorporation of 8-oxoG into DNA during replication (12). *mutM* or *fpg* codes for formamidopyrimidine DNA *N*-glycosylase (Fpg) (13), a protein with *N*-glycosylase and AP lyase activities (14, 15) that acts on both purine and pyrimidine oxidation products (16). However, action on 8-oxoG paired with C is believed to be the main function of Fpg *in vivo.* Fpg activity on 8-oxoG opposite A, a pair that results from replication errors, is greatly reduced (17) to prevent fixation of a G → T transversion. The third gene product in this repair system, MutY, is an *N*-glycosylase that removes A mismatched with 8-oxoG and to a lesser extent with G (18, 19). *E. coli* mutants deficient in any one of these three genes display a mutator phenotype and multiple mutations show synergistic effects (11).

DNA damage-induced mutational spectra obtained from a wide range of species often display regions of high- and low mutation frequency (hot- and coldspots, respectively) which are lesion- and gene-specific (20). These observations imply that a lesion's mutagenic potential is partially determined by the sequence context in which it is presented. Sequence context effects on the efficiency of lesion recognition and removal have been reported for several DNA repair enzymes (21–23). Similarly, a strong influence of neighboring bases on both the efficiency of lesion bypass and the formation and extension of mispairs opposite several DNA lesions has been well documented both *in vitro* (24–27) and *in vivo* (28, 29). We, therefore, set out to define sequence contexts that modulate the mutagenicity of 8-oxoG by affecting the rate at which the lesion is repaired and/or the frequency at which the lesion miscodes for A during DNA synthesis.

METHODS

Chemicals and Enzymes. 8-oxoG phosphoramidite was from Glen Research (Sterling, VA). Wild-type T4 gp43, gp43 deficient in the 3′ → 5′ proofreading exonuclease activity (gp43 exo$^-$), T4 gp44, gp62, gp45, and gp32 were prepared as described (30–32). Fpg protein was overexpressed in *E. coli* and purified following the procedure described for endonuclease VIII (33).

In Vitro **Selection.** A 52-nt oligo (Fig. 1) containing 8-oxoG surrounded by four randomized nucleotides on both the 3′ and 5′ sides was synthesized by standard phosphoramidite chemistry. The 52 mer (5 pmol) was primed with a 5′ end-labeled 16 mer in 10 mM potassium phosphate, 16 mM Tris·HCl (pH 7.4), 65 mM NaCl, 1 mM 2-mercaptoethanol, 150 μM dithiothreiotol, 1 mM EDTA, and 12% glycerol. T4 DNA polymerase holoenzyme complex then was assembled on the oligonucleotide in the absence of Mg^{2+} as described (34). The reaction included 100 μM dNTPs, 1 mM ATP (final concentrations), 140 pmol gp44/62, and 100 pmol gp45 and 10 pmol of gp43. Reactions were initiated by addition of 10 mM $MgCl_2$ (final) and 2.5 nmol of heparin in a total volume of 65 μl and were terminated by a 10-min incubation at 65°C. A molar excess of Fpg over the DNA then was added, and the reaction mix was incubated for 10 min at 37°C. The reaction was stopped by addition of formamide, and the 8-oxoG-containing strands that were refractory to Fpg cleavage were purified by 8 M urea PAGE. As depicted in Fig. 1, the lesion-containing strand was 52 nt long, which was readily separable from the complementary strand (46 nt) and Fpg cleavage product (20 mer). The selected molecules were PCR-amplified and subcloned into pBluescript KS$^+$ by using a restriction endonuclease-free method (35), and the sequence of representative subclones was determined by using standard protocols.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: 8-oxoG, 7,8-dihydro-8-oxoguanine; Fpg, formamidopyrimidine DNA *N*-glycosylase; R, A or G; Y, C or T; K, G or T.
†Present address: Center for Immunology, Washington University School of Medicine, St. Louis, MO 63110.
¶To whom reprint requests should be addressed. e-mail swallace@zoo.uvm.edu.

FIG. 1. Schematic of the *in vitro* selection assay.

Data Analysis. A relational database computer program was used to analyze 76 *in vitro* selected samples. The sequences that represented the original randomized region were searched for all possible strings of two, three, or four nucleotides [both contiguous and interrupted by (n) nucleotides], and statistically under- and overrepresented sequences, as determined by a χ^2 test and P values ≤ 0.05, were isolated.

Fpg Steady-State Kinetics. 8-OxoG was introduced opposite C at one of the position marked 1, 2, or 3 in the 54 mer, 5′-ATTCCAGACTGTCAA**1**AACACGGCG**2**ACCAGTCG-ATCCTGGGCTGCAGGAATTC, 3′-TAAGGTCTGACA-GTTCTTGTGCC**3**CCTGGTCAGCTAGGACCCGACGT-CCTTAAG, or at the position marked 4 in the 24 mer, 5′-GAACTAGTG**4**ATCCCCCGGGCTGC, 3′-CTTGATCAC-CTAGGGGGCCCGACG.

The lesion-containing strand was 5′ end-labeled with ^{32}P, annealed to its complement, adjusted to a range of molar concentrations between 0.5 and 105 nM, and subjected to cleavage by Fpg (final concentration 0.25 nM) at 37°C in 10 mM Tris·HCl (pH 7.5), 1 mM EDTA, and 50 mM NaCl. Following product analysis by denaturing PAGE and quantification by molecular imaging, the reactions' initial rates were determined and plotted as a function of substrate concentration. The resulting curve was fit to a hyperbolic function by using the program SIGMAPLOT, and the values of K_m and k_{cat} were derived.

T4 DNA Polymerase Steady-State Kinetics. The 54 mers containing 8-oxoG at positions 1 and 2 and the 24 mer containing the lesion at position 4 (see above) were used as templates to determine the steady-state kinetics of nucleotide incorporation and extension opposite 8-oxoG by gp43 exo$^-$. Each template was primed with one of three primers, terminating 1 nt 3′ to the lesion (relative to the template strand), terminating with C or terminating with A opposite the lesion. The primed-template concentration was adjusted to 2.5 nM (final during the reaction) in 25 mM Tris·acetate (pH 7.5), 150 mM potassium acetate, and 10 mM 2-mercaptoethanol, and T4 gp 43 exo$^-$ was added (16 nM protein apparent final concentration) in the absence of Mg^{2+}. Reactions were started by addition of 10 mM magnesium acetate and a single dNTP (1 nM to 40 μM final concentration, in ranges appropriate for individual reactions) and were carried out for 5–60 s at 22°C. Single nucleotide incorporation rates were determined as a function of dNTP concentration, and the steady-state rate constants were derived as described for Fpg kinetics. The reported gp43 exo$^-$ final concentration was based on a protein assay and not on titration of the polymerase active site. The concentration of active molecules was likely lower than template concentration because the reaction's initial velocity could be measured in a time scale of 5 to 60 s even at the highest dNTP concentration used [under polymerase excess conditions, DNA synthesis across 8-oxoG has been shown to proceed at a rate of 10s of nucleotides per second (36), making it impossible to measure an initial rate within the time scale used in our experiments].

RESULTS

***In Vitro* Selection Assay.** A 52-nt oligomer containing 8-oxoG in the center of a randomized octamer (Fig. 1) was

Table 1. Observed nucleotide frequency at randomized positions in 76 independent samples of molecules cleaved inefficiently by Fpg

5′ . . . CAGAGT	N	N	N	N	8-oxoG	N	N	N	N	TGACAG . . . 3′
A	38	27	28	24	3	45	30	31	27	
G	10	28	30	26	56	7	21	14	26	
C	17	11	10	14	4	14	11	12	11	
T	11	10	8	12	13	10	14	19	12	
χ^{2*}	36.7	20.0	28.0	10.3	130.7	64.1	14.8	15.1	15.6	
P	<0.001	<0.001	<0.001	<0.025	<0.001	<0.001	<0.005	<0.005	<0.005	

*χ^2 and *P* values are for deviation from a random distribution of 0.25 for each of the four nucleotides.

synthesized chemically and used as template for DNA synthesis by T4 DNA polymerase holoenzyme. The resulting double-stranded molecules then were subjected to exhaustive cleavage by Fpg (see *Methods*), and those representing poor substrates, i.e., molecules refractory to cleavage, were isolated by denaturing PAGE. Less than 1% of the total lesion-containing strands remained uncleaved after the Fpg treatment as determined by phosphorimager analysis (data not shown). The selected molecules were amplified by PCR and subcloned (see Fig. 1 and *Methods*), and the sequences of 76 independent representatives as well as controls that were not subjected to this selection were determined. A highly statistically significant nonrandom sequence distribution was observed readily in the pool of selected molecules (Table 1) but not in the control samples (data not shown). In four of the eight randomized positions (third, second, and first nucleotides to the 5′ and fourth nucleotide to the 3′ side of the lesion), the selected molecules showed a significant ($P < 0.05$) bias for purines, and in the remaining four positions (fourth nucleotide to the 5′ side and first, second, and third nucleotides to the 3′ side of the lesion), there was a significant bias for adenine.

Sequence Analysis of Selected Molecules. The 76 representative samples were analyzed for the presence of consensus sequences. We used a relational database computer program (MICROSOFT FOXPRO) to determine the frequency of every di-, tri-, and tetranucleotide sequence in the sample pool and calculated whether the observed values deviated significantly from what would be expected to arise by chance. Sequence gaps were accommodated to avoid biases toward contiguous sequence contexts (e.g., not only the frequency of the trinucleotide XYZ was determined, but also the frequency of XNYZ, XNNYZ, XYNZ, XYNNZ, etc., in every possible permutation). Statistically over- and underrepresented independent sequences are reported in Table 2. Many additional di- and trinucleotide sequences were overrepresented in the sample pool but might not be independent of the sequences listed in Table 2. For example, there were 11 samples with the sequence 5′-G(N)A (the nucleotide within the parentheses represents the original position of the lesion), a highly overrepresented number ($\chi^2 = 8.2$) in a total sample size of 76. However, there were 45 samples with the sequence 5′-N(N)A, suggesting that 5′-G(N)A was not selected independently of 5′-N(N)A [11 G(N)A/45 N(N)A, $\chi^2 = 0.01$]. Because our selection targeted molecules that were not cleaved after exhaustive treatment with Fpg (Fig. 1), sequences observed in significantly high numbers were taken to represent contexts that reduced the efficiency of Fpg action on 8-oxoG. By inference, sequences observed in significantly low numbers are taken to represent contexts that promoted efficient cleavage of 8-oxoG.

Consensus Sequences that Modulated the Action of Fpg on 8-OxoG. The most striking consensus in sequences promoting inefficient cleavage by Fpg was the dominance of A as the 3′ nearest neighbor of the lesion (59% of all samples). Next, two additional sequences were observed in significantly high numbers, a string of four purines on the 5′ side of the lesion and ARAR on the 3′ side of the lesion. Although samples containing the consensus 5′-(N)ARAR are a subset of those containing 5′-(N)A, their observed frequency in the pool of the latter sequence (12/45) was much higher than expected by random distribution ($\chi^2 = 30$), suggesting that they were selected independently. On the other hand, 5′-(N)A was still highly overrepresented in samples not containing 5′-(N)ARAR (33/64, $\chi^2 = 18.1$). It is therefore probable that the presence of A as the 3′ nearest neighbor exerted a negative effect on 8-oxoG removal by Fpg, whereas the presence of 5′-(N)ARAR enhanced this effect. Similarly, the presence of four purines on the 5′ side of the lesion appeared to enhance the effect of the 3′-A [nine samples carrying 5′-RRRR(N)A of 45 with 5′-(N)A, $\chi^2 = 13.6$]. However, there were eight samples with the sequence 5′-RRRR(N) in 31 samples not containing a 3′-A ($\chi^2 = 19.0$), suggesting that the presence of four purines on the 5′ side of the lesion exerted some negative effect on lesion removal independently of the 3′ nearest neighbor. Furthermore, of 17 samples containing the consensus 5′-RRRR(N) and 12 samples containing 5′-(N)ARAR, only two samples contained both, suggesting that the two sequences were selected independently of each other and that either consensus was sufficient to effect poor lesion removal by Fpg.

Three dinucleotide sequences were underrepresented in the pool of selected samples (Table 2). These contexts were present in significantly low numbers even when the observed low frequency of pyrimidines in the total sample pool (Table 1) was taken into account. For example, the sequence 5′-(N)YC was not only significantly underrepresented in 76 total samples ($\chi^2 = 7.6$, Table 2) but also in 24 samples containing the sequence 5′-(N)Y ($\chi^2 = 4.17$) and 11 samples with the sequence 5′-(N)NC ($\chi^2 = 3.68$). We take this to indicate that the underrepresented sequences are contexts that promoted efficient cleavage of the lesion by Fpg, more so than generally pyrimidine-rich sequences.

Table 2. *In vitro*-selected sequence contexts for inefficient and efficient cleavage of 8-oxoG by Fpg

	Consensus	*O*	*E*	χ^2	*P*
Inefficient cleavage consensus,	5′ (N)A	45	19	35.6	<0.001
n = 76	5′ (N)ARAR	12	1.2	98.5	<0.001
	5′RRRR (N)	17	4.8	31.6	<0.001
	5′RRRR (N)A	9	1.2	51.4	<0.001
Efficient cleavage consensus,	5′ (N)YC	1	9.5	7.6	<0.01
n = 76	5′ Y(N)K	4	19	11.8	<0.001
	5′ YY(N)	4	19	11.8	<0.001

Listed sequences are for the strand that originally contained the lesion (N). Following convention, R = purine, Y = pyrimidine, and K = G or T.

Table 3. Statistical analysis of the appearance of the sequence 5′G(T)A in the pool of *in vitro*-selected samples

	O	*E*	χ^2	*P*
5′-GTA/5′-G(N)A (n = 11)	7	2.8	6.6	<0.01
5′-GTA/5′-N(T)N (n = 13)	7	0.8	47.1	<0.001

Reported sequences are for the template strand and the original position of the lesion is indicated by parentheses.

Consensus Sequences that Modulated the Frequency of Misincorporating A Opposite 8-OxoG by DNA Polymerase. A very interesting observation came from a small subset of samples (17%) in which T rather than G was recovered at the original position of the lesion. As alluded to earlier, these samples may define a consensus sequence that promoted simultaneously inefficient cleavage of 8-oxoG by Fpg and a high frequency of misincorporating A opposite the lesion by DNA polymerase. Such a sequence context constitutes an ideal putative G → T transversion hotspot. Although it is clear that the transversions observed in our sample pool were generated by *Taq* DNA polymerase during PCR amplification, contexts that had directed the misincorporation of A opposite the lesion during second strand synthesis by T4 DNA polymerase already would have been selected during the Fpg cleavage step of the assay [8-oxoG paired with A is a poor substrate of Fpg (17)]. In addition, we were interested in identifying contexts where *Taq* polymerase misinserted A opposite the lesion because such contexts could have similar effects on all DNA polymerases. Of 13 samples in which T was observed at the original position of the lesion, 10 had A as the 3′ nearest neighbor (χ^2 = 14), 9 had G as the 5′ nearest neighbor (χ^2 = 10.2), and 7 had both (χ^2 = 47.1; Table 3). No significant consensus was observed at other positions in the randomized portion of the 52 mer substrate, suggesting that the sequence 5′-G(N)A was sufficient to effect a higher rate of miscoding for A by 8-oxoG. Conversely, in 11 samples in which G and A were observed as, respectively, the 5′ and 3′ nearest neighbors (regardless of the nucleotide recovered at the position of the lesion), 7 samples contained T at the position of the lesion, again a significantly overrepresented number (Table 3 reports a χ^2 of 6.6 for 7 observed/11 total based on equal probability for all four nucleotides, which is an underestimate in view of the fact that G would be expected to predominate). Because of the limited pool size (76 samples), we cannot exclude the presence of longer contexts with higher miscoding potential.

Verification of the *in Vitro* Selection Assay. To verify the effectiveness of the selection scheme, we determined the kinetic parameters for action of Fpg on 54 nucleotide oligomers containing 8-oxoG in the contexts 5′-CC(8-oxoG)CC, 5′-AA(8-oxoG)AA, and 5′-TG(8-oxoG)AC (Methods and Table 4). The first sequence was predicted to promote efficient lesion removal, whereas the latter two were predicted to promote the opposite effect (Table 2). A 24 mer containing the lesion in the context 5′-TG(8-oxoG)AT also was used. As indicated by the k_{cat}/K_m values (which provide an estimate of a reaction's efficiency), Fpg removed 8-oxoG from the 5′-CC(8-oxoG)CC context 9-fold more efficiently than from the 5′-AA(8-oxoG)AA context, in full agreement with our predictions. More interesting, the putative hotspot context 5′-G(8-oxoG)A lowered the efficiency of Fpg action by 33-fold compared with the CC(8-oxoG)CC context. The hotspot contexts promoted inefficient removal of the lesion predominantly by raising the K_m of the reaction.

Similarly, we verified the elevated miscoding potential of the lesion in the 5′-G(8-oxoG)A contexts by measuring the kinetic parameters for T4 DNA polymerase synthesis past 8-oxoG in three of the four contexts described in Table 4. Parameters for incorporation of C and misincorporation of A opposite 8-oxoG as well as extension of the two resulting pairs were determined. (Values for a 3′ → 5′ exonuclease-free polymerase are shown; similar observations were made with the proofreading-proficient polymerase). Confirming predictions made in Table 3, A was incorporated opposite the lesion 33-fold more efficiently in the context 5′-G(8-oxoG)A than the context 5′-CC(8-oxoG)CC and ≈4-fold more than in the context 5′-AA(8-oxoG)AA. It should be noted that the efficiency of incorporating C opposite the lesion was also higher in the 5′-G(8-oxoG)A context (14- and 4-fold, respectively) than 5′-CC(8-oxoG)CC and 5′-AA(8-oxoG)AA. As for the Fpg

Table 4. Kinetics of Fpg and T4 DNA polymerase processing of 8-oxoG within putative mutational hot- and coldspots

	5′-CC(8-oxoG)CC	5′-AA(8-oxoG)AA	5′-TG(8-oxoG)AC	5′-TG(8-oxoG)AT
Sequence context				
Fpg cleavage				
K_m, nM	1.3 ± 0.17	17.8 ± 2.15	43.8 ± 9.15	44.4 ± 7.55
k_{cat}, min^{-1}	6.0 ± 2.16	9.6 ± 1.16	7.8 ± 1.23	6.0 ± 0.83
k_{cat}/K_m, $(nM \cdot min)^{-1}$	4.62	0.54	0.18	0.14
Synthesis by T4 DNA Pol (3′ → 5′ exonuclease-free)				
Insertion of C opposite 8-oxoG				
K_m, μM	0.683 ± 0.094	0.332 ± 0.032	ND	0.11 ± 0.008
V_{max}, $nM_{DNA} \cdot min^{-1}$	0.35 ± 0.006	0.59 ± 0.017	ND	0.8 ± 0.014
V_{max}/K_m, $nM_{DNA} \cdot (\mu M_{dNTP} \cdot min)^{-1}$	0.51	1.78	ND	7.3
Insertion of A opposite 8-oxoG				
K_m, μM	8.87 ± 0.45	1.38 ± 0.07	ND	0.34 ± 0.024
V_{max}, $nM_{DNA} \cdot min^{-1}$	0.66 ± 0.01	0.86 ± 0.014	ND	0.78 ± 0.014
V_{max}/K_m, $nM_{DNA} \cdot (\mu M_{dNTP} \cdot min)^{-1}$	0.07	0.62	ND	2.3
Extension of C:8-oxoG pair				
K_m, μM	8.24 ± 0.56	1.01 ± 0.217	ND	0.911 ± 0.23
V_{max}, $nM_{DNA} \cdot min^{-1}$	0.74 ± 0.015	0.72 ± 0.049	ND	0.5 ± 0.048
V_{max}/K_m, $nM_{DNA} \cdot (\mu M_{dNTP} \cdot min)^{-1}$	0.09	0.71	ND	0.55
Extension of A:8-oxoG pair				
K_m, μM	0.18 ± 0.006	0.102 ± 0.01	ND	0.32 ± 0.021
V_{max}, $nM_{DNA} \cdot min^{-1}$	0.81 ± 0.002	0.72 ± 0.016	ND	0.66 ± 0.01
V_{max}/K_m, $nM_{DNA} \cdot (\mu M_{dNTP} \cdot min)^{-1}$	4.5	7.1	ND	2.1
$(V_{max}/K_m)_{inc} \cdot (V_{max}/K_m)_{ext}$				
C:8-oxoG pair	0.05	1.26		4.02
A:8-oxoG pair	0.32	4.4		4.83

ND, not determined.

Table 5. Nearest neighbors frequencies at sites of spontaneous, ionizing radiation-, or oxidation-induced G → T transversions in the *lacI* gene of *E. coli*

Nearest neighbors distribution at G → T transversion sites in *E. coli lacI*								
	Total number of mutations (n = 147)				Number of mutated sites (n = 36)			
Sequence context	O	E	χ^2	P	O	E	χ^2	P
5′-N(G → T)A-3′	100	37	109	<0.001	19	9	11	<0.001
5′-G(G → T)A-3′	42	9	117	<0.001	7	2	10	<0.005
5′-C(G → T)C-3′	11	9	0.4	<0.5	2	2	0	0
5′-T(G → T)T-3′	2	9	5.6	<0.01	1	2	0.7	<0.25
Nearest neighbors distribution at G sites in *E. coli lacI* (n = 362)								
					O	E	χ^2	P
5′-N(G)A-3′					68	91	5.6	<0.025
5′-G(G)A-3′					19	23	0.6	<0.25
5′-C(G)C-3′					38	23	10.45	<0.005
5′-T(G)T-3′					19	23	0.6	<0.25

cleavage reaction, sequence context effects were observed predominantly as differences in K_m values.

For translesion synthesis to be successful, efficient extension of the nucleotide incorporated opposite the lesion is required. Earlier work has shown that, for several lesions, including 8-oxoG and several DNA polymerases, incorporation of a mutagenic nucleotide is less efficient than the nonmutagenic nucleotide (A and C, respectively for 8-oxoG). But curiously, extension of the mutagenic pair is more efficient than the nonmutagenic one (10, 37). Data presented in Table 4 confirm these observations. In all three sequence contexts, extension of the 8-oxoG:A pair was 4- to 50-fold more efficient than extension of the 8-oxoG:C pair. Of interest, although the extension of the A mispair was 50-fold more efficient than formation of the mispair in the 5′-CC(8-oxoG)CC context (V_{max}/K_m values of 4.5 and 0.07, respectively), the two steps were equally efficient in the context 5′-G(8-oxoG)A (values of 2.3 and 2.1).

When the kinetics of the combined incorporation/extension reactions are compared, the 5′-G(8-oxoG)A context promoted nonmutagenic translesion synthesis 80-fold more efficiently than the 5′-CC(8-oxoG)CC context and promoted mutagenic translesion synthesis 15-fold more efficiently. By further combining the effects on Fpg and DNA polymerase, the sequence context 5′-G(8-oxoG)A was shown to have the potential of generating G → T transversions 450-fold more efficiently than the context 5′-CC(8-oxoG)CC.

Correlation of the *in Vitro* Derived Consensus Sequences with *in Vivo* Mutational Spectra. The authenticity of our *in vitro*-selected putative hot- and coldspots could be further tested by searching *in vivo* mutational spectra. We searched a database of mutations scored in the *E. coli lacI* gene (38), which is sufficiently large in total number of mutations and represents a relatively large sequence diversity, for all spontaneous or free radical-induced G → T transversions and determined the frequency of the 16 possible sequence contexts 5′ N(mutation)N (reference to G → T transversions throughout this report is restricted to mutations generated spontaneously or by free radical-producing agents). As can be seen in Table 5, a striking 100/147 total mutations had A as the 3′ nearest neighbor (χ^2 = 109), and 42 of the latter had G as the 5′ nearest neighbor (χ^2 = 117). Although it was clear that the sequence 5′-G(G → T)A was overrepresented in the context 5′-N(G → T)A (42/100, χ^2 = 11.6), the presence of 3′A was significant even at sites where the 5′ nearest neighbor was not G (58/105, χ^2 = 38.4). Because these mutations were scored based on a phenotypic change, the data can be skewed by the presence of a large number of mutations at one site. We therefore searched for the frequency of G → T transversions as a function of sequence context without considering the number of mutations at each site. Of 37 different sites where G → Ts were observed, 19 had A as the 3′ nearest neighbor and 7 fell in the sequence 5′-G(G → T)A, both highly overrepresented frequencies. These observations are in striking agreement with the predictions of our *in vitro* selection. Furthermore, only 11 mutations at 2 sites were scored where C was the nearest neighbor on both sides of the mutations, and 2 mutations at one site were surrounded by Ts (representing random distribution in the first case and significant underrepresentation in the second). Because the *lacI* gene encompasses a limited sequence diversity, it was possible that the observed biases toward G → T transversions in the context 5′-G(G)A simply reflected a higher incidence of this sequence in the gene. The lower portion of Table 5 rules out this possibility. There are 362 Gs in *lacI*, and only 68 have A as the 3′ nearest neighbor, significantly fewer than expected by random distribution (χ^2 = 5.6). Similarly, 19 Gs fall in the context 5′-G(G)A, reflecting random distribution. On the other hand, 38 Gs are in the context 5′ C(G)C, significantly more than expected by chance (χ^2 = 10.45).

Although we used prokaryotic proteins in our selection, we were interested to see whether this putative mutational hotspot is significantly overrepresented in spectra of human genes. In the tumor suppresser p53 gene, G → T transversions are found within the context 5′-G(G)A in 60/672 total mutations at 12/104 sites (38) (both numbers significantly overrepresented, χ^2 = 7.71, 4.65, respectively). Similarly, G → T transversions in the human factor IX gene (antihemophilic factor B) have a significantly overrepresented 10/71 mutations in the putative hotspot context scored at 8/47 mutable sites (χ^2 = 6.97, 8.72, respectively).

DISCUSSION

Based on the biological observation that mutations cluster in hot and coldspots, we initiated this project on the premise that sequence context strongly modulates the mutagenic outcome of an important base oxidation product, 8-oxoG. We went on to isolate such sequences *in vitro* and to verify their competence by kinetic analysis. Finally, we were able to go full circle and find very strong correlation between our data and available *in vivo* mutational spectra, demonstrating the effectiveness of the *in vitro* selection technique.

As for most other DNA *N*-glycosylases, Fpg action is limited to removal of lesions in double-stranded DNA (39). The randomized portion in the oligonucleotide substrate used in the selection assay necessitated the use of DNA polymerase to generate a perfect complement for the chemically synthesized single strand containing 8-oxoG (Fig. 1). An interesting and advantageous outcome of this step was the generation of

double-stranded molecules in which the DNA polymerase, due to sequence context, incorporated A preferentially opposite the lesion. Because Fpg acts inefficiently on 8-oxoG paired with A (17), our selection scheme was predicted to identify sequence contexts that rendered 8-oxoG a poor substrate of Fpg either by reducing the efficiency of removing lesions paired with C or by increasing the efficiency of misincorporating A opposite the lesion. Most interesting, our results showed that, within the putative hotspot 5′-G(8-oxoG)A, both mechanisms operated.

The 5′-G(8-oxoG)A putative hotspot not only supported higher efficiency of misincorporating A opposite the lesion by T4 DNA polymerase relative to other contexts (see *Results* and Table 4) but also higher efficiency of incorporating C, suggesting that this context had the general effect of minimizing lesion-induced distortion in the DNA structure. In addition, T4 DNA polymerase incorporated A opposite 8-oxoG and extended the resulting primer terminus with almost identical efficiency (Table 4) in this sequence context, simulating natural DNA synthesis (the rates of successive polymerization steps remain relatively constant during DNA synthesis on undamaged templates). In contrast, large variations in the efficiency of these two steps were observed during the incorporation of C opposite the lesion in the 5′-G(8-oxoG)A context and the incorporation of A opposite the lesion in the 5′-CC(8-oxoG)CC context. The latter observation suggested that the putative hotspot reduced lesion-induced distortions in DNA structure more efficiently when A was paired with the lesion.

The strong correlation between predictions made by the *in vitro* selection and bacterial *in vivo* spectra on the one hand and human *in vivo* spectra on the other suggested that the sequence context effect was mediated by altering lesion presentation independently of bound proteins (repair glycosylases or DNA polymerases). On the other hand, the stronger correlation between the *in vitro* data and the bacterial mutational spectrum (relative to the human spectra) can be interpreted to reflect differences in the prokaryotic and eukaryotic enzymes involved in processing the lesion. Although it is probable that both notions are correct, further work is needed to resolve the relative effect of this sequence context on different polymerases and glycosylases. In addition, correlation with *in vivo* data must take into account the influence of the putative hotspot sequence context on MutY.

We are currently in the process of further verifying our results by introducing plasmids carrying 8-oxoG in the putative hot- and coldspots into bacterial and mammalian cells and measuring the rate of G → T transversions in the different contexts. We are also using purine analogs as nearest neighbors of 8-oxoG to define the exocyclic groups involved in effecting the higher mutagenic potential of the lesion. Finally, sequence context effects on modulating the mutagenic potential of other oxidative lesions such as 5-hydroxycytosine will be addressed using a similar *in vitro* selection scheme.

We are grateful to our colleagues Drs. Robert Melamede and Wah Kow for stimulating discussions; Dr. Lois E. Rabow for providing Fpg during the preliminary stages of the selection assay; and Ms. Pam Vacek for assistance in the statistical analyses. This work was supported by National Institutes of Health Grants CA 33657 and CA 52040 awarded by the National Cancer Institute.

1. Denq, R. Y. & Fridovich, I. (1989) *Free Radical Biol. Med.* **6,** 123–129.
2. Adelman, R., Saul, R. L. & Ames, B. N. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2706–2708.
3. Ames, B. N. (1989) *Mutat. Res.* **214,** 41–46.
4. Cadet, J. & Berger, M. (1985) *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* **47,** 127–143.
5. Fraga, C. G., Shigenaga, M. K., Park, J. W., Degan, P. & Ames, B. N. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4533–4537.
6. Kamiya, H., Murata-Kamiya, N., Koizume, S., Inoue, H., Nishimura, S. & Ohtsuka, E. (1995) *Carcinogenesis* **16,** 883–889.
7. Michaels, M. L. & Miller, J. H. (1992) *J. Bacteriol.* **174,** 6321–6325.
8. Moriya, M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1122–1126.
9. Wood, M. L., Esteve, A., Morningstar, M. L., Kuziemko, G. M. & Essigmann, J. M. (1992) *Nucleic Acids Res.* **20,** 6023–6032.
10. Shibutani, S., Takeshita, M. & Grollman, A. P. (1991) *Nature (London)* **349,** 431–434.
11. Tajiri, T., Maki, H. & Sekiguchi, M. (1995) *Mutat. Res.* **336,** 257–267.
12. Maki, H. & Sekiguchi, M. (1992) *Nature (London)* **355,** 273–275.
13. Michaels, M. L., Pham, L., Cruz, C. & Miller, J. H. (1991) *Nucleic Acids Res.* **19,** 3629–3632.
14. Boiteux, S., Gajewski, E., Laval, J. & Dizdaroglu, M. (1992) *Biochemistry* **31,** 106–110.
15. Tchou, J. & Grollman, A. P. (1995) *J. Biol. Chem.* **270,** 11671–11677.
16. Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. S. (1994) *J. Biol. Chem.* **269,** 18814–18820.
17. Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P. & Johnson, F. (1994) *J. Biol. Chem.* **269,** 15318–15324.
18. Michaels, M. L., Tchou, J., Grollman, A. P. & Miller, J. H. (1992) *Biochemistry* **31,** 10964–10968.
19. Tsai-Wu, J. J., Liu, H. F. & Lu, A. L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8779–8783.
20. Friedberg, E. C., Walker, G. C. & Siede, W. (1995) *DNA Repair and Mutagenesis* (Am. Soc. Microbiol., Washington, DC).
21. Nilsen, H., Yazdankhah, S. P., Eftedal, I. & Krokan, H. E. (1995) *FEBS Lett.* **362,** 205–209.
22. Mu, D., Bertrand-Burggraf, E., Huang, J. C., Fuchs, R. P., Sancar, A. & Fuchs, R. P. (1994) *Nucleic Acids Res.* **22,** 4869–4871.
23. Svoboda, D. L., Smith, C. A., Taylor, J. S. & Sancar, A. (1993) *J. Biol. Chem.* **268,** 10694–10700.
24. Belguise-Valladier, P. & Fuchs, R. P. (1995) *J. Mol. Biol.* **249,** 903–913.
25. Cai, H., Bloom, L. B., Eritja, R. & Goodman, M. F. (1993) *J. Biol. Chem.* **268,** 23567–23572.
26. Dosanjh, M. K., Galeros, G., Goodman, M. F. & Singer, B. (1991) *Biochemistry* **30,** 11595–11599.
27. Shibutani, S. & Grollman, A. P. (1993) *J. Biol. Chem.* **268,** 11703–11710.
28. Evans, J., Maccabee, M., Hatahet, Z., Courcelle, J., Bockrath, R., Ide, H. & Wallace, S. (1993) *Mutat. Res.* **299,** 147–156.
29. Maccabee, M., Evans, J. S., Glackin, M. P., Hatahet, Z. & Wallace, S. S. (1994) *J. Mol. Biol.* **236,** 514–530.
30. Reha-Krantz, L. J. & Nonay, R. L. (1993) *J. Biol. Chem.* **268,** 27100–27108.
31. Bittner, M., Burke, R. L. & Alberts, B. M. (1979) *J. Biol. Chem.* **254,** 9565–9572.
32. Morris, C.-F., Hama-Inaba, H., Mace, D., Sinha, N. K. & Alberts, B. (1979) *J. Biol. Chem.* **254,** 6787–6796.
33. Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. & Wallace, S. S. (1997) *J. Bacteriol.* **179,** 3773–3782.
34. Reddy, M. K., Weitzel, S. E. & von Hippel, P. H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 3211–3215.
35. Zhou, M. & Hatahet, Z. (1995) *Nucleic Acids Res.* **23,** 1089–1090.
36. Furge, L. L. & Guengerich, F. P. (1997) *Biochemistry* **36,** 6475–6487.
37. Kamiya, H., Murata-Kamiya, N., Fujimuro, M., Kido, K., Inoue, H., Nishimura, S., Masutani, C., Hanaoka, F. & Ohtsuka, E. (1995) *Jpn. J. Cancer Res.* **86,** 270–276.
38. Hutchinson, F. & Donnellan, J. E. J. (1997) *Nucleic Acids Res.* **25,** 192–195.
39. Tchou, J., Kasai, H., Shibutani, S., Chung, M. H., Laval, J., Grollman, A. P. & Nishimura, S. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4690–4694.